

COMPONENTS OF A DIGITAL LIBRARY

Dr. Sachin Y. Vaidya

Librarian

Dr. Haribhau Admane Arts Commerce
College, Saoner

vidya.s30@gmail.com

Abstract:

Digital libraries promise new societal benefits, especially for e-learning in digital or mobile times, starting with the elimination of the time and space constraints of traditional bricks-and-mortar libraries. The library and information professionals are required to acquire such knowledge and skills as the library is one of the highly IT influenced service profession. This paper gives an overview of current trends in digital library research consists of digital library characteristic, advantage, disadvantages and function. This paper also highlights on the impact of information technology on the traditional library.

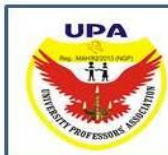
Introduction:

Establishing digital library resources and services require a great deal of infrastructural components that are not available off-the-shelf as packaged solution. There is no turn-key, monolithic systems available for digital libraries, instead digital libraries are collection of disparate systems and resources connected through a network and made interoperable using open system architecture and open protocol and are integrated within one interface, currently the web interface. Use of open architecture and open standards make it possible that pieces of required infrastructure, be it hardware, software or accessories, are gathered from different vendors in the marketplace and integrated to construct a working digital library environment. Several components required for establishing a digital library are internal to the institutions, but several others are distributed across the Internet, owned and controlled by a large number of independent players. The task of building a digital library, therefore, requires a great deal of integration of various components (Flecker, D., 2001). Major components required for a digital library can broadly be divided into six major categories.

These components are described briefly in this module. However, separate modules are devoted to impart detailed information on each of the six components of digital library mentioned above.

Collections Infrastructure:

The most important component of a digital library is the digital collection it holds or has access to. Viability and extent of the usefulness of a digital library depends upon the critical mass of digital collection it has. The collection infrastructure typically consists of two components, i.e. metadata and digital objects that a digital library holds. The metadata provides bibliographic or index information for the digital objects. While digital objects are the primary documents that users are interested to access, it is metadata that facilitates their identification, retrieval and location using variety of search techniques. Information



content of a digital library, depending on the media type it contains, may include a combination of structured / unstructured text, numerical data, scanned images, graphics, audio and video recordings and other multimedia content. Different types of resources need to be handled differently in a digital library. The libraries, irrespective of media types that they house, i.e. print, audio-visual or digital, are primarily responsible for identifying, selecting, organizing, preserving and providing access to diverse categories of resources to their users. The transition from traditional library to digital library cannot happen overnight in a single step, rather this transition is gradual and incremental in nature. As such, the traditional libraries are not becoming digital libraries, but are increasingly acquiring access to ever growing digital collections for their users either by licensing of e-resources available in the market place or by its acquisition on one-time purchase and perpetual access basis. Collections in digital libraries may also consist of datasets that are “borne digital” or existing printed documents converted into digital format through scanning. Creating virtual libraries, library portals or subject gateways are also considered as an important digital library collection. Collection management in a digital or hybrid library need to have pre-defined policies and practices similar to those being followed in traditional library while keeping in view the issues and complexities that are especially related to digital materials. The current electronic publishing market consists of traditional players such as commercial publishers, scholarly societies, university presses offering electronic versions of their print journals as well as several new enterprises offering new products and services that are “borne digital”. The market also has several aggregators that provide electronic resources in a given discipline sourced from different publishers. These publishers offer a variety of electronic resources including electronic journals, electronic books, conference proceedings, online courseware, learning materials, tutorials, guides, manuals, patents, standards, electronic e-prints (preprints and postprints), technical reports, electronic theses and dissertations, online databases and databanks, dictionaries, encyclopedias, subject portals or pathfinders. Major

publishers, besides offering their electronic journals are now offering electronic books either directly through their Web sites or in partnership with other publishers or through aggregators like e-brary, NetLibrary, Questia, 24x7, Knovel, etc. Moreover, more than 32,000 books are available free of cost through Project Gutenberg. These electronic resources are available on variable pricing model.

Digital Knowledge Organization:

Traditional library consists of physical objects such as books, journals, conference documents, standards, patents, video, microfilms and CDs that are organized into various collections such as Text Books, General Books, Reference Books, Rare Books, Audio-visuals, CD ROM Collections and Journals. Each collection is further organized using classification schemes such as Dewey Decimal Classification, Library of Congress Classification, Universal Decimal Classification, Colon Classification, etc. so as to bring books on same subject together and facilitate browsing documents on the shelves. Moreover, each book is catalogued and assigned subject headings using standard subject headings and thesauri like Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH), Sear's Subject Headings, etc. so as to facilitate their retrieval using Library OPAC. While physical libraries are organized at physical level, i.e. books, journals, theses, reports, reference books, textbooks, etc., digital libraries are organized at digital objects level which may include a combination of structured / unstructured text, numeric data, scanned images, graphics, articles in a journal or chapters in a book and other multimedia objects.



Access Infrastructure:

Browse, Search and Navigation Interfaces of Digital Library An effective and efficient access mechanism that allows a user to browse, search and navigate digital resources becomes necessary as electronic resources of a collection grow in number and complexion. While the access infrastructure for a traditional library is OPAC/WebOPAC (including journals holding), the access infrastructure for digital libraries consists of browse, search and navigational interfaces for individual digital libraries, specialized indices for specialized local collections, portals or subject gateways for web resources and an integrated interface for all e-resources accessible to a given library including library OPAC.

Network and Computing Infrastructure:

A typical digital library in a distributed client-server environment consists of hardware and software components at server side as well as at the client's side. Clients are machines that are used for accessing digital library by users while the server hosts databases, digital objects, browse, search and navigational interfaces to facilitate its access.

Computer hardware, software and network infrastructure for a digital library can broadly be divided into the following four categories:

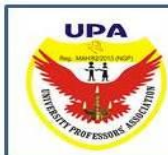
- i) Server-side Hardware Components including input devices, storage devices, Communication Devices, etc.;
- ii) Server-side Software Components including image capturing or scanning software, image enhancement and manipulation software, web servers, information retrieval software, Optical Character Recognition (OCR) software, Database Management System (BDMS) Software, Digital Rights Management (DRM), etc.;

Intellectual Property Rights (IPR):

Copyright has been called the "single most vexing barrier to digital library development" (Chepesuk, 1997). The current paper-based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed, easily copied, and remotely accessible by multiple users simultaneously. The libraries, unlike private businesses or publishers that own their information, are simply caretakers of the information.

Physical ownership or possession of material by a library is not necessarily an indicator of ownership of corresponding copyright. It is unlikely that libraries will ever be able to freely digitize and provide access to the copyrighted materials in their collections. Instead, the developers of digital libraries are obliged to take permission for inclusion of copyrighted material in digital form or develop mechanisms for managing copyright, mechanisms that allow them to provide information without violating copyright. Copyrights and IPR issues are governed by the constitutions of various countries and through international treaties like the Berne Convention.

"Fair Use" is an exception to copyright protection that permits limited use of copyrighted material without explicit permission of the owner for non-commercial and non-



profiteducational purposes. Protection and ownership of intellectual property in the age ofelectronic information are especially confusing in light of traditional copyright laws.

Discussions are taking place at various platforms to review the existing copyright laws in the light of electronic information. Since the images are electronically forwarded around theInternet, it becomes very difficult to control and define what can and cannot be done. Copyright is manifested in terms of licenses and agreements in the digital world. A library isrequired to sign licenses to acquire access to a digital collection. The terms of licenses fordigital collection varies in terms of conditions, the variety of pricing models and accesslimitations(see Collection Development – licensing contents). The library associations andpublishers are working on model licenses that can be adopted uniformly. The libraries cannegotiate with the publishers on behalf of their institutions or as a consortium of libraries.

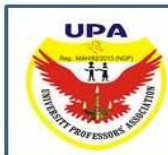
User Authentication:

A combination of one or more of the authentication mechanisms are deployed by thepublishers for allowing access to the digital content to the authorized users hosted in digitallibraries. These authentication mechanisms are: i) Log-in ID and Password-based Access; ii)IP Filtering; iii) Web Cookies; iv) Web Proxy;v) Athens; vi) Shibboleth; and vii)Referring URL. These authentication mechanisms are described in detail in the module onaccess management.

User Authorization:

The process of authentication ascertains the identity of a user, while authorization defines hisor her permissions in terms of access to e-resources and extent of its usage. Authorization isgranted to the successfully authenticate users according to his / her rights informationavailable in the Access Management System (AMS). A user duly authenticated by one of theauthentication mechanism described above may actually be entitled to access only a portionof digital collection subscribed by his / her institution. For example, an authenticated usermay be authorised to access electronic journals from a publisher's web site but not electronicbooks, reference sources or other resources depending on what his institution has subscribedto. Typically, all users in an institution are authorized to access all the subscribed e-resources.However, it is possible to define different levels of authorization for different categories ofpersonnel in an institution. Besides, authorizing users of a digital collection, authorizationalso addresses the issue of responsibilities assigned to different personnel involved in thedevelopment of a digital library and their respective authorities in terms of addition, deletion,editing and uploading of records into a digital library. Personnel involved in the developmentof a digital library may be assigned different levels of authority. Authorization is morechallenging than authentication, especially for widely distributed digital libraries. Access control is one method for enforcing authorization. Typically, it assumes that the user or entityhas already been authenticated. Access control policies that are in vogue include i)Mandatory Access Control (MAC); ii) Discretionary Access Control (DAC); iii) Role BasedAccess Control (RBAC); and iv) Content Dependent Access Control (CDAC).These accesscontrol policies are described in detail in the module on access management.

Digital Library Services:



The library research and development in digital libraries, in the beginning, was focused mainly towards providing search and browsing interface to its collection. However, providing access to its resources is only one of the several services offered by a traditional library to its users. Reference services, for example, provide personalized services to a user with human touch. The importance of reference service has increased many-fold with introduction of new information technologies in libraries. Users, who are not well versed with use of web and Internet technology, find it difficult to retrieve information from plethora of resources accessible to them from various digital repositories. Sloan (1998) emphasized that technology and information sources, on its own, cannot make up an effective digital library. Helping users in finding resources, either in physical or electronic environment, is the foremost task of a librarian.

E-mail Alerts:

The service, variably called as E-mail Alert, Table of Contents Alert, News Alert, etc., offer the ability to set up an e-mail alerts for the table of contents from a specific journal or group of journals by the end user. A user can subscribe to e-mail alerts to get periodic emails with links to new content automatically that are added to the publisher's web site. The service, offered by most of the digital libraries and databases, can broadly be equated to Current Awareness Services (CAS) offered by traditional libraries.

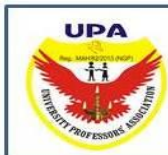
Web Feeds: RSS Feeds or Atom:

Web feeds are data formats used for providing users with frequently updated content. The two main web feed formats are RSS and Atom. RSS stands for Real Simple Syndication or Rich Site Summary and Atom format was developed as an alternative to RSS. The technology, on one hand allows a web site to list the newest published updates (like table of contents of journals, new articles) through a technology called XML, on the other hand, it

facilitates a web user to keep track of new updates on chosen website(s). Like a personal search assistant, RSS feed readers visit pre-defined web sites, look for updated information and fetch it automatically on to the user's desktop. In order to use RSS Feed, users are required to download RSS feed reader or RSS feed aggregator, which can be web-based, desktop-based, or mobile-device-based and then "subscribe" to the RSS feeds by copying a link from the web site of a digital repository into their feed reader. The reader can then check the subscribed feeds to see if any of those feeds have new content since the last time it was checked, and if so, retrieve that content and present it to the user. Both RSS and Atom are supported by most of the feed readers.

Electronic Document Delivery Services:

The term "electronic document delivery systems" implies delivery of the electronic version of a document that might involve reproduction of an electronic copy of a document if it is not already available in electronic format. However, with availability of most of the peer-reviewed research journals in electronic format, most publishers and aggregators facilitate online electronic document delivery services that allow a user to download an article in full-text from their site at a pre-determined cost. Different publishers and aggregators offer different payment options, i.e. some charge each time the journal is used, whereas others provide restriction-free access for an annual subscription.



References:

- Arora, Jagdish. Building digital libraries: An overview. *DESIDOC Bulletin of Information Technology*, 21(6), 3-24, 2001.
- Chepesuk, R. The future is here: America's libraries go digital. *American Libraries*, 2(1),47-49, 1997.
- Cox, John E. (1997). Publishers, publishing and the Internet: how journal publishing will survive and prosper in the electronic age. *Electronic Library*, 15(2), 125-131.
- Digital Millennium Copyright Act (1998) of USA(<http://www.copyright.gov/legislation/dmca.pdf>) (Last visited on 30th January, 2013)
- Flecker, D. Preserving scholarly e-journals. *D-Lib Magazine*, 7(9), 2001. (<http://www.dlib.org/dlib/september01/flecker/09flecker.html>)
- Marchionini, G., Plaisant, C. and Komlodi, A. Interfaces and tools for the Library of Congress National Digital Library Program. *Information Processing and Management*,34(5), 535-555, 1998.
- Networked Digital Libraries of Theses and dissertations (NDLTD) project(<http://www.ndltd.org/>). (Last visited on 25th Decemeber, 2023)
- Sloan, Bernard G. Services perspectives for the digital library: Remote reference services. *Library Trends*, 47(2), 1998.